Towards TEI Lex-0 Publisher

Wolfgang Meier

Magdalena Turska

Dark side of the TEI

- Overwhelming vocabulary of more than 600 elements

- High variance in encoded material with no best practice consensus

- Only covers the text encoding, agnostic about broader context

Requirements

integrate all editorial schools
address all humanities disciplines
cover all textual genres and kinds of objects
explicit semantic
machine processable and human readable
subject oriented - data instead of media
single source - multiple publications

Power to the editor!

TEI Simple rationale

Relief to the developer

… and even funding agencies should like it

projects are **unique**

technical requirements **repeat**

The usual menu:

- browse, sort

- facets, filters, full text search

- document transformation into multiple outputs: HTML website, print-ready PDF, ePub…

- facsimile viewer, map visualisation…

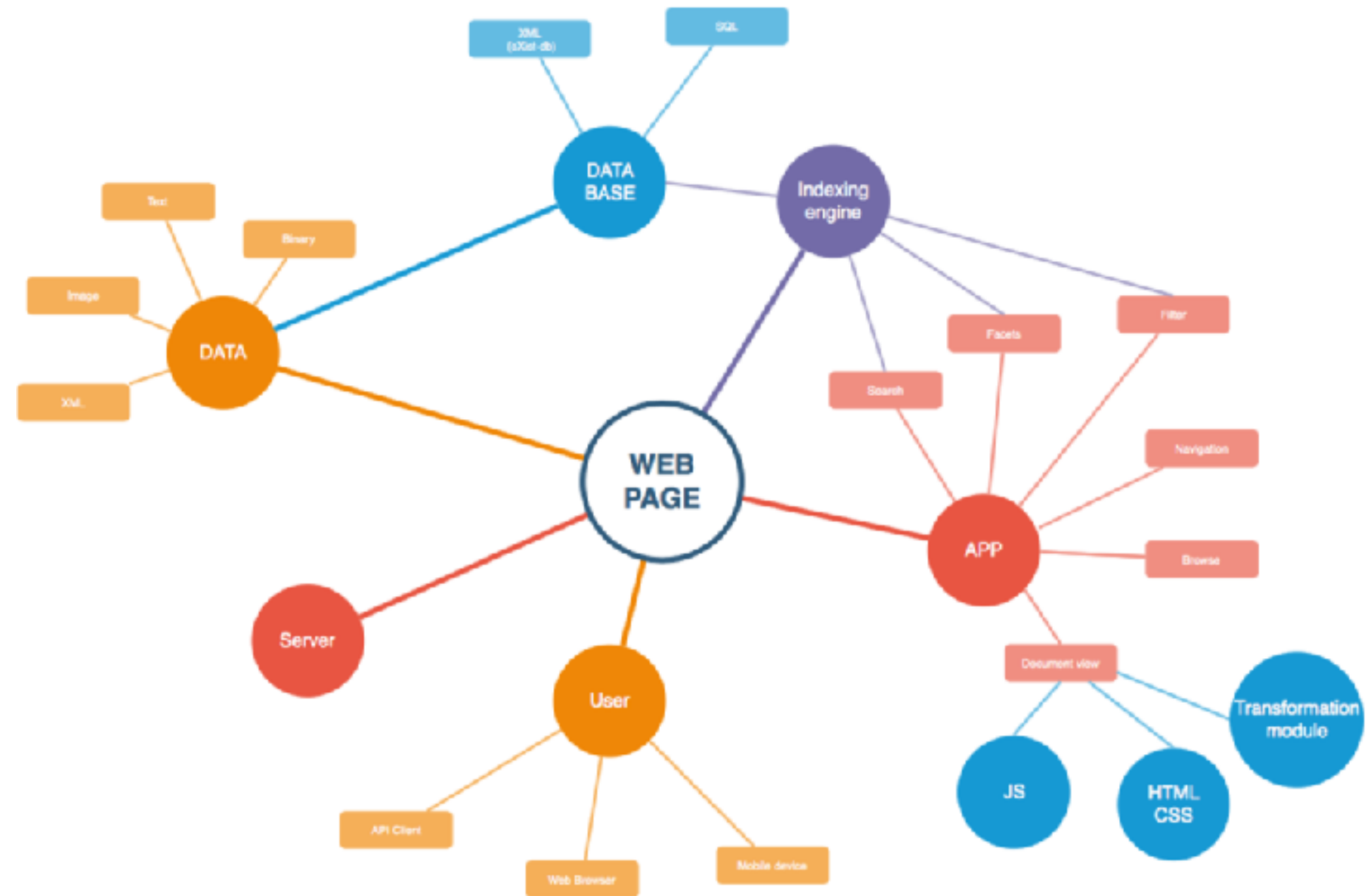# Can't you just do it in Word?

LaTeX?
WordPress?
surely Python! ;-)

Components behind a web page

Sustainability

- Data sources are standardized and reusable …

- … but the software used for their publication is not!

Use case: Transformation module

- Write an XSLT?

- Learn XQuery?

- Fight with XSL-FO or dig up LaTeX again?

What are the prospects?

- a few thousand lines of code for transformation to HTML …

- … and yet another few thousands for PDF

- maintenance becomes horror!

Abstraction to the rescue:

TEI Processing Model

ODD = One Document Does it all

- With TEI processing model the promise was finally met!

- One document covers schema definition, documentation and intended appearance of TEI texts

- Simple, standardised syntax, will be understood in years to come

- Following the principle of literate programming: complete documentation in TEI

What is the Processing Model?

- Media-agnostic description of output transformations

- Maps TEI elements to abstract behaviours

- Expresses the intended processing for a document within the TEI vocabulary itself

What does it gain?

- just 3 new elements and 24 predefined behaviours

- Easy to tweak and maintain models

- Just one ODD for all output formats

- Order of magnitude code base reduction ~300 lines vs ~3k

TEI Source + ODD
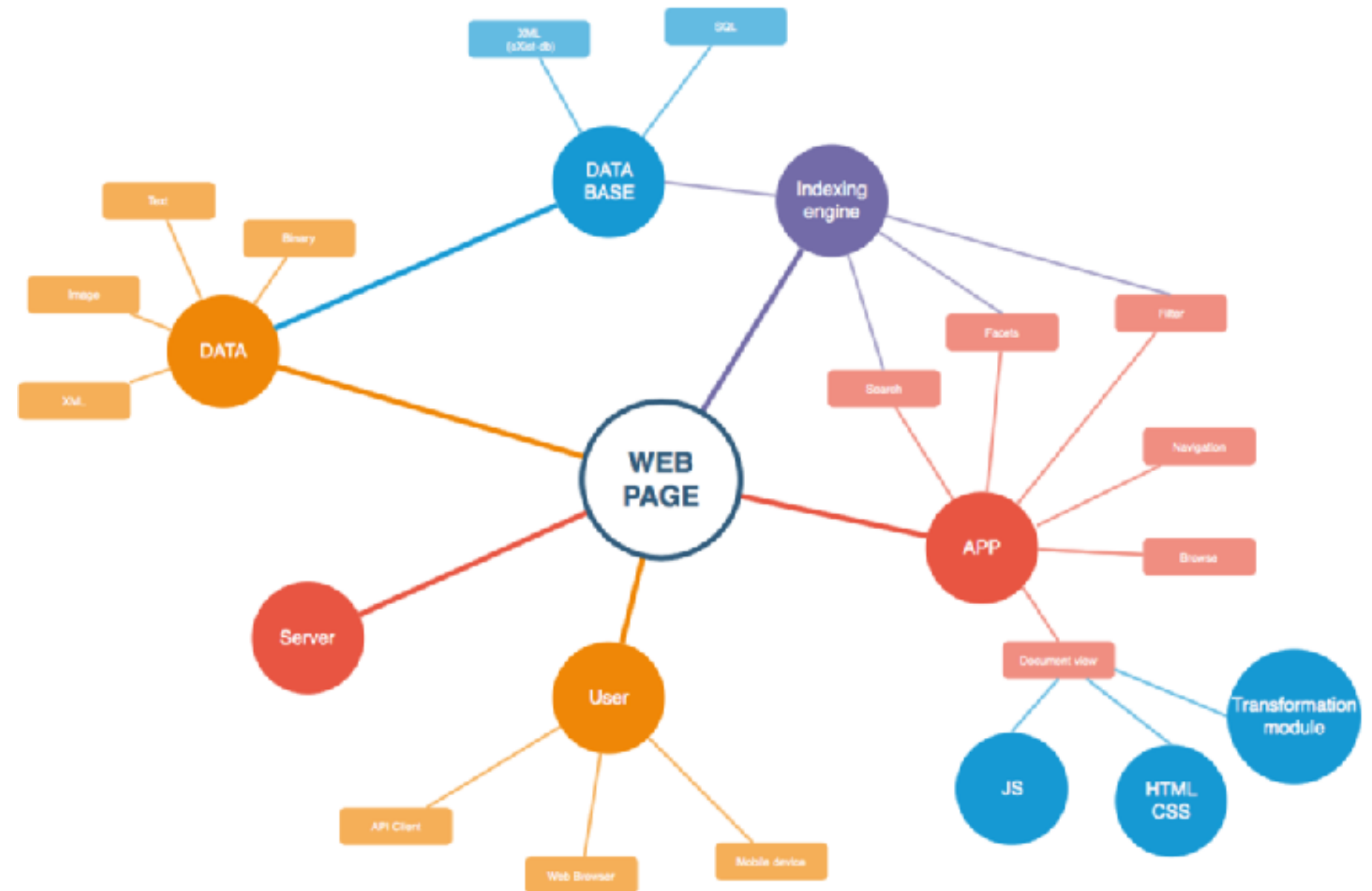## enough to reconstruct an edition

Dodis Demo

vs

Quaderni di Dodis

Put it to the test

- Various edition projects we realized prove:

  - powerful enough to cover **complex transformation** needs

  - a truly universal tool for **any** kind of edition

  - efficient and fast: **war tested** on large sites

  - suitable for **any** XML, not just TEI

repeat
for other
parts of the
Project

tei
Publisher

TEI Publisher

- started as a reference implementation of the TEI processing model in 2015

- soon adding a customisable application and app generator

- before switching to web component based modular design in 2018

Vision

- Open-source tool for publishing digital editions

- Sustainable and interoperable thanks to standard-conformance and modular approach

- All editions can communicate

- As simple as possible but allowing for any complexity where required

Community Project

- **TEI Publisher** and **eXist-db** are open source, thus *free software* - in the sense of free speech, not free beer - it still requires resources to develop and maintain

- User contribution in various forms is **essential**

- Open Source First - when the goal is to create a **universal**, **open source** tool, all contributing projects have to be willing to accept the overhead this means

TEI Publisher +
Web components

All functionality as small lego blocks to be
freely arranged

Web Components

- Self-contained, encapsulated custom HTML elements

- Part of W3C HTML5 standard

- Native implementations coming to most browsers

Inter
oper
ability

e, Sunday, 29 September 1872.

e, Saturday, 26 October 1872.

e, Friday, 13 December 1872.

e, mid-January 1873.

e, Tuesday, 28 January 1873.

e, Monday, 17 March 1873.

e, Monday, 24 March 1873.

Language

- [ ] en   10
- [ ] fr   322
- [ ] nl   595

From

- [ ] Jo van Gogh-Bonger   5
- [ ] Joseph Roulin   4
- [ ] Paul Gauguin   16
- [ ] Theo van Gogh   39
- [ ] Vincent van Gogh   844

To

- [ ] Anthon van Rappard   58
- [ ] Emile Bernard   22
- [ ] Theo van Gogh   654
- [ ] Vincent van Gogh   84
- [ ] Willemien van Gogh   22

Date

- [ ] 1882   101
- [ ] 1883   123
- [ ] 1888   153
- [ ] 1889   107
- [ ] 1890   77

Declaration	1
Interview	1
Journal	1
Letter	3
Memo	21
Minutes	2
Political report	2
Report	2
Telegram	30

Language ☐ Show all

English	17
German	17
Hebrew	6
Polish	7
Turkish	6

Author ☐ Show all

Delworth, Thomas W. (1929–2012)	3
Mekik, Metin (1932–2013)	4
Navon, Benjamin (1933–)	4
Shevardnadze, Eduard (1928–2014)	2

Draft memo by the Director of the 4th Department of the Polish Mi

Memorandum Concerning the Citizens of the German Democratic Republic who Rem

Thursday, 14th September, 1989

Memo by the Austrian Ministry of Foreign Affairs

Das Gespenst der deutschen Wiedervereinigung

Tuesday, 19th September, 1989

Telegram from the Israeli Ambassador in Bonn, Benjamin Navon, to

Friday, 27th October, 1989

Telegram from the Dutch Ambassador in East Berlin, Egbert Jacobs

GDR-Internally

Wednesday, 8th November, 1989

Extract from the Diary of the Protocol Officer at the Embassy of the

Kohl-Besuch und Mauerfall

Thursday, 9th November, 1989

What is eXist?

eXist Core Features

1. Accept any XML of any size or complexity

2. Directly locate any node in a huge collection

3. Avoid loading documents into memory

4. Evaluate XPath/XQuery via indexes, not tree traversals

1. Accept any XML

- XML should be designed to match the domain it describes, not the system to process it!
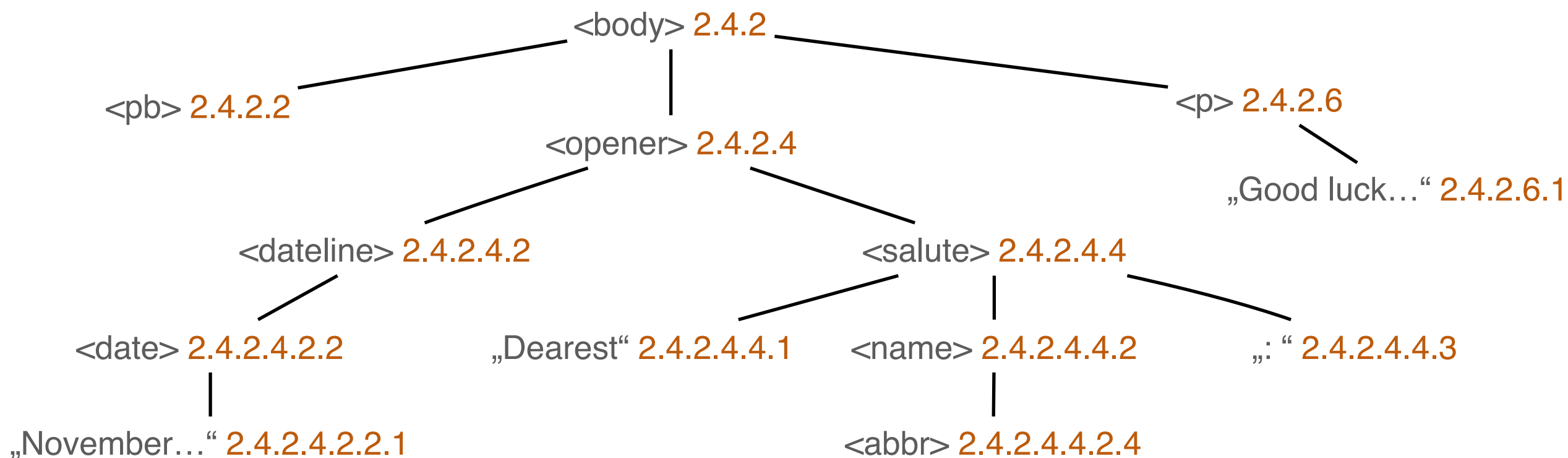
- Size of a single document should be irrelevant

2. Directly locate any node in a collection

- Every node in eXist has a unique address

- Immediately locate any node, no matter how deep it is nested …

- … or how large its parent document is

```xml
<text exist:id="2.4" xmlns:exist="http://exist.sourceforge.net/NS/exist">
    <body exist:id="2.4.2">
        <pb exist:id="2.4.2.2" facs="WG.57.11.15.a.jpg"/>
        <opener exist:id="2.4.2.4">
            <dateline exist:id="2.4.2.4.2">
                <date exist:id="2.4.2.4.2.2" when="1957-11-15">November 15, 1957</date>
            </dateline>
            <salute exist:id="2.4.2.4.4"> Dearest <name exist:id="2.4.2.4.4.2" ref="#GravesWilliam" type="person">
                <abbr exist:id="2.4.2.4.4.2.4">Wm</abbr>
            </name>: </salute>
        </opener>
        <p exist:id="2.4.2.6"> Good luck in your interview. If you are wholly at your ease - and why not? - all
            will go well. But try to raise some sort of enthusiasm for your proposed career:
            dont-care-ism doesn't go down well. </p>
    </body>
</text>
```

<body> 2.4.2

<pb> 2.4.2.2

<p> 2.4.2.6

<opener> 2.4.2.4

„Good luck…" 2.4.2.6.1

<dateline> 2.4.2.4.2

<salute> 2.4.2.4.4

<date> 2.4.2.4.2.2

„Dearest" 2.4.2.4.4.1

<name> 2.4.2.4.4.2

„: " 2.4.2.4.4.3

„November…" 2.4.2.4.2.2.1

<abbr> 2.4.2.4.4.2.4

3. Avoid loading documents into memory

- eXist mostly operates with node ids, not the actual nodes

- Documents or fragments are never loaded into memory - unless user explicitly asks for it

- Avoid access to the actual node stored on disk

4. Evaluate XPath/XQuery via indexes

- XPath describes traversing a tree

- eXist uses indexes to take shortcuts everywhere

- Can determine the relationship between nodes based on node identifiers

- Uses fast set operations to compare node identifiers instead of traversing actual document tree

More than an XML Database

- eXist aims to provide an entire ecosystem for developing XML-based applications

- All our applications are written in XQuery: no other language needed

- Standardized packaging for apps and libraries

- Install apps with a mouse click

TEI Processing Model

behaviour

common concepts inspired by natural language

block

inline

with formatting

link

footnote

Where?
What?
When?
Which?
Why?

model

where, to which element

what behaviour to apply

when is this model applicable, in what context

which parameters to pass

why so? document!

```
<elementSpec ident="name" mode="change">
    <model behaviour="inline">
        <outputRendition>
            color: #CC6600;
        </outputRendition>
    </model>
</elementSpec>
```

@ident: name of the element this rule applies to

@behaviour=inline: process as an inline element

<outputRendition>: output style given as CSS

^ name + 🗑 📋

^ model [inline] ↓ ↑ 🗑 ⧉   Output: ⇕

Description: [Document the model]

Behaviour: inline

Predicate: [XPath condition: model applies only if matched]

CSS Class: [Define CSS class name (for external CSS)]

Parameters   +

Renditions   +   ☐ Use source rendition

Scope: 🗑

color: #CC6600;

```
<elementSpec ident="name" mode="change">
    <model predicate="@type='place'" behaviour="inline">
        <outputRendition>
            font-variant: small-caps;
        </outputRendition>
    </model>
    <model behaviour="inline">
        <outputRendition>
            color: #CC6600;
        </outputRendition>
    </model>
</elementSpec>
```

- a predicate defines a condition: model only applied to matching elements

- takes XPath expression using current element as context

∧ name + 🗑 📋

∧ model [inline] ↓ ↑ 🗑 ⧉   Output: ⬍

   [@type="place"]

Description: [Document the model]

Behaviour: inline

Predicate: @type="place"

CSS Class: [Define CSS class name (for external CSS)]

   Parameters   +

   Renditions   +   ☐ Use source rendition

   Scope: 🗑

      font-variant: small-caps;

∨ model [inline] ↓ ↑ 🗑 ⧉   Output: ⬍

```
<elementSpec ident="date" mode="change">
    <model output="web" predicate="@when" behaviour="alternate">
        <param name="default" value="."/>
        <param name="alternate" value="format-date(@when, '[FNn], [D1o] [MNn], [Y]')"/>
    </model>
    <model predicate="text()" behaviour="inline"/>
</elementSpec>
```

- Behaviours can take parameters

- e.g. „alternate" requires parameters „default" and „alternate"

- Parameter „content" is predefined: overwrite it to define what should be processed by the model

ODD „Inheritance"

- ODDs can extend or overwrite a parent ODD

- examples extend teipublisher.odd which inherits from tei_simplePrint.odd

- tei_simplePrint.odd defines standard processing rules for a subset of TEI

- visual editor automatically inserts existing rules when you overwrite an element spec

∧ **date** + 🗑 📋

∧ model [alternate] ↓ ↑ 🗑 ⧉   Output: web ↕

[@when]

Description: [Document the model]

Behaviour: alternate

Predicate: @when

CSS Class: [Define CSS class name (for external CSS)]

Parameters +

default   .

alternate   `format-date(@when, '[FNn], [D1o] [MNn], [Y]')`   🗑

Renditions +   ☐ Use source rendition

∨ model [inline] ↓ ↑ 🗑 ⧉   Output: ↕

[text()]

```
<elementSpec ident="date" mode="change">
    <modelSequence predicate="@when" behaviour="inline">
        <model behaviour="inline"/>
        <model behaviour="note">
            <param name="content"
                value="format-date(@when, '[FNn], [D1o] [MNn], [Y]')"/>
        </model>
    </modelSequence>
    <model behaviour="inline"/>
</elementSpec>
```

- <modelSequence> applies all child models in sequence

- without it, only the first model would be applied

Document view

XML source

     + ODD

   + HTML page template

         with Web Components

Hernán CORTÉS to Ioannes DANTISCUS, Mexico 1531-08-01

Dos Cartas De Vuestra Señoría, una^BCz, 1596, p duplicada de otra, he rreceuido, y en cada una dellas los versos que Vuestra Señoría hizo intitulados a la magestad cesárea, y leilos muchas vezes por bien comprehenderlos. Pareciéronme en su brevedad y mucha sentencia bien ser hechos de la mano de Vuestra Señoría, porque de otra no era bastante. Beso las manos de Vuestra Señoría por enbiarmelos y por sus cartas, pues dellas entendí tener salud Vuestra Magnífica Persona, y por otra de Gerónimo Sailer que Vuestra Señoría havía obispado, de que holgue mucho; plega a Nuestro Señor que por pontificado se rrenuncie que bien creerá Vuestra Señoría que no me pesaría dello.

Yo rrecibo tanta merced con las cartas de Vuestra Señoría que querría que, aunque la distancia del largo camino estorva mi deseo, se diese orden como se hiziese lo posible. Y pues Gerónimo rreside en la corte de Castilla, por la via suya podrá Vuestra Señoría escrevirme. Suplico a Vuestra Señoría lo haga y me escriva siempre de su salud y prosperidad y de todo lo que sucediere en la estada de la magestad cesárea allá, asi en lo

Otrzymałem od Waszej Wielmożności dwa listy, jeden jest kopią drugiego, a w każdym z nich wiersze, które Wasza Wielmożność napisała, zadedykowane Cesarzowi. Czytałem je wielokrotnie, aby je dobrze zrozumieć. Ich zwięzłość i głębokie znaczenie poświadczają, że są dziełem Waszej Wielmożności, ponieważ kto inny by tak nie potrafił. Całuję ręce Waszej Wielmożności za posłanie mi ich i za Jej listy, z których dowiedziałem się, że Wasza Dostojność ma się zdrowo, a z innego listu, od Hieronima Sailera, że Wasza Wielmożność otrzymała biskupstwo, z czego bardzo się cieszę. Błagaj Pana Boga, żeby cię uchronił przed godnością papieską — co Waszej Wielmożności łatwo może przyjść na myśl — żebym potem tego nie żałował.

Listy Waszej Wielmożności to dla mnie taka łaska, że chciałbym aby powtarzała się stale, na ile to możliwe, chociaż na przeszkodzie mojemu pragnieniu stoi wielkość dzielącej nas drogi. A ponieważ Hieronim przebywa na Dworze Kastylijskim, za jego pośrednictwem może Wasza

Component communication channels

```
<body unresolved="unresolved" data-template="i18n:translate" data-template-catalogues="resources/i18n">
    <pb-page data-template="pages:pb-page">
        <pb-document id="document1" data-template="pages:pb-document"/>
        <app-drawer-layout force-narrow="force-narrow">
            <app-drawer data-template="templates:include" data-template-path="templates/drawer.html"/>
            <app-header-layout>
                <app-header slot="header" reveals="reveals" fixed="fixed" effects="waterfall" data-template="browse:fix-links">
                    <app-toolbar data-template="templates:include" data-template-path="templates/menu.html"/>
                    <app-toolbar data-template="templates:include" data-template-path="templates/toolbar.html"/>
                </app-header>

                <section class="breadcrumbs">
                    <pb-view id="title-view1" src="document1" xpath="//teiHeader" view="single">
                        <pb-param name="mode" value="title"/>
                    </pb-view>
                </section>
                <main class="content body">
                    <pb-view id="view1" src="document1" xpath="//text[@type='source']" column-separator=".tei-cb"
                        append-footnotes="append-footnotes"
                        subscribe="transcription" emit="transcription" wait-for="#facsimile"/>
                    <pb-facsimile id="facsimile" base-uri="https://apps.existsolutions.com/cantaloupe/iiif/2/"
                        default-zoom-level="0" show-navigation-control="show-navigation-control"
                        show-navigator="show-navigator" subscribe="transcription"/>
                    <pb-view id="view2" src="document1" map="cortez-translation" xpath="//text[@type='source']"
                        subscribe="transcription" column-separator=".tei-cb" append-footnotes="append-footnotes"/>
                </main>
            </app-header-layout>
        </app-drawer-layout>
    </pb-page>
</body>
```

templates/cortez.html

templates/toolbar.html

```
<app-toolbar xmlns:i18n="http://exist-db.org/xquery/i18n" class="toolbar" sticky="sticky">
    <paper-icon-button id="tocToggle" class="toc-toggle" icon="icons:view-list"/>
    <pb-zoom emit="transcription" direction="in" icon="icons:zoom-in"/>
    <pb-zoom emit="transcription" direction="out" icon="icons:zoom-out"/>
    <pb-navigation emit="transcription" keyboard="left" subscribe="transcription" direction="backward" unit="page">
        <paper-icon-button icon="icons:chevron-left"/>
    </pb-navigation>
    <pb-navigation emit="transcription" keyboard="right" subscribe="transcription" direction="forward" unit="page">
        <paper-icon-button icon="icons:chevron-right"/>
    </pb-navigation>
    <paper-icon-button icon="icons:menu" drawer-toggle="drawer-toggle"/>
    <pb-progress subscribe="transcription" indeterminate="indeterminate" bottom-item="bottom-item"/>
</app-toolbar>
```

Open Source First

- TEI Publisher is free software in the sense of free speech, not free beer

- User contribution is not optional, but essential for this kind of community project

Wanted

- Documentation and teaching material authors

- TEI examples to showcase specific domains

- Presentation templates based on examples other users might reuse

- Code contributions

- Cross-funding, project cooperations, workshops

Reach out to us:

- teipublisher.com

- https://gitlab.existsolutions.com/tei-publisher

- Slack